Gates & Cooper LLP

Howard Hughes Center
6701 Center Drive West, Suite 1050
Los Angeles, California 90045

FAX TRANSMISSION TO USPTO

TO: Commissioner for Patents
Attn: Examiner Michael B. Holmes
Patent Examining Corps
Facsimile Center
Alexandria, VA 22313-1450

FROM: George H. Gates
OUR REF.: 8224
TELEPHONE: (310) 642-4146

Total pages, including cover letter: 35

PTO FAX NUMBER: 571-273-8300

If you do NOT receive all of the pages, please telephone us at (310) 641-8797, or fax us at (310) 641-8798.

Title of Document Transmitted:	**TRANSMITTAL SHEETS AND BRIEF OF APPELLANTS.**
Applicant:	Timothy E. Miller et al.
Serial No.:	09/806,743
Filed:	April 2, 2001
Group Art Unit:	2121
Title:	DATA MINING ASSISTS IN A RELATIONAL DATABASE MANAGEMENT SYSTEM
Our Ref. No.:	8224

Please charge all fees to Deposit Account No. 14-0225 of NCR Corporation, the assignee of the present application.

By: [signature]
Name: George H. Gates
Reg. No.: 33,500

I hereby certify that this paper is being transmitted by facsimile to the U.S. Patent and Trademark Office on the date shown below.

Barbara Senty
Signature

September 6, 2005
Date

GHG/bjs

G&C 30145.315-US-WO

Due Date: September 7, 2005

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

Applicant:	Timothy E. Miller et al.	Examiner:	Michael B. Holmes
Serial No.:	09/806,743	Group Art Unit:	2121
Filed:	April 2, 2001	Docket:	8224
Title:	DATA MINING ASSISTS IN A RELATIONAL DATABASE MANAGEMENT SYSTEM		

CERTIFICATE OF MAILING OR TRANSMISSION UNDER 37 CFR 1.8

I hereby certify that this correspondence is being filed via facsimile transmission to the U.S. Patent and Trademark Office on September 6, 2005.

By:

Name: George H. Gates

MAIL STOP APPEAL BRIEF - PATENTS
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

We are transmitting herewith the attached:

- ☒ Transmittal sheet, in duplicate, containing a Certificate of Mailing or Transmission under 37 CFR 1.8.
- ☒ Brief of Appellant(s).
- ☒ Charge the Fee for the Brief of Appellant(s) in the amount of $500.00 to the Deposit Account.

Please consider this a PETITION FOR EXTENSION OF TIME for a sufficient number of months to enter these papers, if appropriate.

Please charge all fees to Deposit Account No: 14-0225 of NCR Corporation (the assignee of the present application). A duplicate of this paper is enclosed.

Customer Number 22462
GATES & COOPER LLP
Howard Hughes Center
6701 Center Drive West, Suite 1050
Los Angeles, CA 90045
(310) 641-8797

By:
Name: George H. Gates
Reg. No.: 33,500
GHG/bjs

G&C 30145.315-US-WO

Due Date: September 7, 2005

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE
BEFORE THE BOARD OF PATENT APPEALS AND INTERFERENCES

In re Application of:	
Inventor: Timothy E. Miller et al.	Examiner: Michael B. Holmes
Serial #: 09/806,743	Group Art Unit: 2121
Filed: April 2, 2001	Appeal No.: _________
Title: DATA MINING ASSISTS IN A RELATIONAL DATABASE MANAGEMENT SYSTEM	

BRIEF OF APPELLANTS

MAIL STOP APPEAL BRIEF - PATENTS
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

In accordance with 37 CFR §41.37, Appellants' attorney submits herwith the Brief of Appellants' on appeal from the final rejection in the above-identified application, as set forth in the Office Action dated April 8, 2005.

Please charge the amount of $500 to cover the required fee for filing this Brief as set forth under 37 CFR §41.37(a)(2) and 37 CFR §41.20(b)(2) to Deposit Account No. 14-0225 of NCR Corporation, the assignee of the present application.

Also, please charge any additional fees or credit any overpayments to Deposit Account No. 14-0225 of NCR Corporation.

I. REAL PARTY IN INTEREST

The real party in interest is NCR Corporation, the assignee of the present application.

II. RELATED APPEALS AND INTERFERENCES

There was a related appeal for Application Serial No. 09/410,528, filed on October 1, 1999, by Brian D. Tate, James E. Pricer, Tej Anand, and Randy G. Kerber, entitled SQL-Based Analytic Algorithm for Association, attorney's docket number 8219, now U.S. Patent No. 6,611,829, issued August 26, 2003.

There is a related appeal for Application Serial No. 09/410,531, filed on October 1, 1999, by James D. Hildreth, entitled SQL-Based Analytic Algorithm for Clustering, attorney's docket number 8220.

There is a related appeal for Application Serial No. 09,410,530, filed on October 1, 1999, by Todd M. Brye, entitled SQL-Based Analytic Algorithm for Rule Induction, attorney's docket number 8221.

III. STATUS OF CLAIMS

Claims 1-23 were cancelled.

Claims 24-83 are pending in the application.

Claims 24-34, 36-40, 42, 43, 46-55, 57-61, 63-74, 76-80, 82 and 83 were rejected under 35 U.S.C. §102(e) as being anticipated by Iyer et al. (Iyer), U.S. Patent No. 5,899,992.

Claims 35, 56, and 75 were rejected under 35 U.S.C. §103(a) as being unpatentable over Iyer in view of SAS Institute Inc., SAS OnlineDoc, Version 8, Cary, NC: SAS Institute Inc., (09/1999).

Claims 41, 62 and 81 were rejected under 35 U.S.C. §103(a) as being unpatentable over Iyer in view of SPRINT: A Scalable Parallel Classifier for Data Mining, John Shafer, Rakesh Agrawal, Manish Mehta, Proceeding of the 22nd VLDB Conference Mumbai (Bombay), India, 1996.

Claims 44 and 45 were rejected under 35 U.S.C. §103(a) as being unpatentable over Iyer in view of Bridges, U.S. Patent No. 5,548,770.

Claims 24-83 are being appealed.

IV. STATUS OF AMENDMENTS

An amendment under 37 C.F.R. §1.116 was submitted on September 13, 2004 in response to the final Office Action dated July 12, 2004, wherein claims 33, 39, 52, 54, 56, 58, 60, 71, 73, 75, 77 and 79 were amended.

An Advisory Action mailed October 13, 2004 did not indicate that the amendments were being entered for the purposes of appeal.

Record is made of a telephone interview between the below-signed Attorney and Examiner Holmes that took place on November 12, 2004, wherein the below-signed Attorney inquired as to the status of the claim amendments.

Thereafter, a Supplemental Advisory Action was mailed on November 23, 2004, which included an Interview Summary, indicating that the amendments were being entered for the purposes of appeal.

A Notice of Appeal was filed on November 12, 2004.

An Appeal Brief was filed on January 12, 2005.

An Office Action was mailed on April 8, 2005.

A Notice of Appeal was filed on July 7, 2005.

V. SUMMARY OF THE INVENTION

Appellants' independent claims 24, 44 and 45 are generally directed to computer-implemented system for performing data mining applications.

Claim 24 recites a computer-implemented system for performing data mining applications. The system includes a computer having one or more data storage devices connected thereto, wherein a relational database is stored on one or more of the data storage devices. The system also includes a relational database management system, executed by the computer, for accessing the relational database stored on the data storage devices. Finally, the system includes an analytic application programming interface (API) that generates a set of scalable data mining functions including queries for execution by the relational database management system, executed by the computer, for performing data mining operations directly within the database management system.

Claim 44 recites a method for performing data mining applications. The method includes the steps of storing a relational database on one or more data storage devices connected to a computer, accessing the relational database stored on the data storage devices using a relational database management system, and executing an analytic application programming interface (API) that generates a set of scalable data mining functions including queries for execution by the relational database management system, for performing data mining operations directly within the database management system.

Claim 45 recites an article of manufacture comprising logic embodying a method for performing data mining applications. The method includes the steps of storing a relational database on one or more data storage devices connected to a computer, accessing the relational database stored on the data storage devices using a relational database management system, and executing an analytic application programming interface (API) that generates a set of scalable data mining functions including queries for execution by the relational database management system, for performing data mining operations directly within the database management system.

With regard to the claims, Appellants' attorney requests that the Board refer to the specification generally. Specific portions of the specification that directly relate to the claims on appeal include:

(a) at page 6, line 6 through page 6, line 26, and in FIG. 1 as reference numbers 100, 102, 104, 106, 108, 110, 112, 114, 116 and 118;

(b) at page 7, lines 5-9;

(c) at page 7, lines 13-18, and in FIG. 2 as reference numbers 202 and 204;

(d) at page 8, lines 11-15, and in FIGS. 1 and 2 as reference numbers 110, 114, 118, 200, 202 and 204;

(e) at page 9, line 14 through page 15, line 16; and

(f) at page 16, line 33 through page 18, line 15, in FIGS. 3, 4 and 5 as reference numbers 300-308, 400-410 and 500-512.

VI. <u>GROUNDS OF REJECTION TO BE REVIEWED ON APPEAL</u>

1. Whether claims 24-34, 36-40, 42, 43, 46-55, 57-61, 63-74, 76-80, 83 and 83 are anticipated under 35 U.S.C. §102(e) by Iyer, U.S. Patent No. 5,899,992.

2. Whether claims 35, 56, and 75 are obvious under 35 U.S.C. §103(a) over Iyer in view of SAS Institute Inc., SAS OnlineDoc, Version 8, Cary, NC: SAS Institute Inc., (SAS).

3. Whether claims 41, 62, and 81 are obvious under 35 U.S.C. §103(a) over Iyer in view of SPRINT: A Scalable Parallel Classifier for Data Mining, John Shafer, Rakesh Agrawal, Manish Mehta, Proceeding of the 22nd VLDB Conference Mumbai (Bombay), India, 1996.

4. Whether claims 44 and 45 are obvious under 35 U.S.C. §103(a) as being unpatentable over Iyer in view of Bridges, U.S. Patent No. 5,548,770.

VII. GROUPING OF CLAIMS

The rejected claims do not stand or fall together. Instead, the claims are grouped as follows:

(a) claims 24, 26, 42, 44, 45, 47, 63, 66, 82 stand or fall together,

(b) claims 25, 46 and 65 stand or fall together,

(c) claims 27, 48 and 67 stand or fall together,

(d) claims 28-31, 35-41, 49-52, 56-62, 68-71, 75-81 stand or fall together,

(e) claims 32, 53 and 72 stand or fall together,

(f) claims 33, 54 and 73 stand or fall together,

(g) claims 34, 55 and 74 stand or fall together, and

(h) claims 43, 64 and 83 stand or fall together.

Separate arguments for each of the groups of claims are provided below.

VIII. ARGUMENT

A. The Office Action Rejections

In paragraph (6) of the Office Action, claims 24-34, 36-40, 42, 43, 46-55, 57-61, 63-74, 76-80, 82 and 83 were rejected under 35 U.S.C. §102(e) as being anticipated by Iyer, U.S. Patent No. 5,899,992. In paragraphs (9) and (10) of the Office Action, claims 35, 56, and 75 were rejected under 35 U.S.C. §103(a) as being unpatentable over Iyer in view of SAS Institute Inc., SAS OnlineDoc, Version 8, Cary, NC: SAS Institute Inc., (SAS). In paragraphs (11) and (12) of the Office Action, claims 41, 62, and 81 were rejected under 35 U.S.C. §103(a) as being unpatentable over Iyer in view of SPRINT: A Scalable Parallel Classifier for Data Mining, John

Shafer, Rakesh Agrawal, Manish Mehta, Proceeding of the 22nd VLDB Conference Mumbai (Bombay), India, 1996. In paragraphs (13) and (14) of the Office Action, claims 44 and 45 were rejected under 35 U.S.C. §103(a) as being unpatentable over Iyer in view of Bridges, U.S. Patent No. 5,548,770.

Appellants' attorney respectfully traverses these rejections.

B. The Appellants' Independent Claims

Appellants' independent claims 24, 44 and 45 are generally directed to computer-implemented system for performing data mining applications. Claim 1 is representative and comprises the elements of:

(a) a computer having one or more data storage devices connected thereto, wherein a relational database is stored on one or more of the data storage devices;

(b) a relational database management system, executed by the computer, for accessing the relational database stored on the data storage devices; and

(c) an analytic application programming interface (API) that generates a set of scalable data mining functions including queries for execution by the relational database management system, executed by the computer, for performing data mining operations directly within the database management system.

C. The Iyer Reference

Iyer describes a method, apparatus, and article of manufacture for a computer implemented scaleable set-oriented classifier. The scalable set-oriented classifier stores set-oriented data as a table in a relational database. The table is comprised of rows having attributes. The scalable set-oriented classifier classifies the rows by building a classification tree. The scalable set-oriented classifier determines a gini index value for each split value of each attribute for each node that can be partitioned in the classification tree. The scalable set-oriented classifier selects an attribute and a split value for each node that can be partitioned based on the determined gini index value corresponding to the split value. Then, the scalable set-oriented classifier grows the classification tree by another level based on the selected attribute and split

value for each node. The scalable set-oriented classifier repeats this process until each row of the table has been classified in the classification tree.

D. The SAS Reference

SAS describes a correlation matrix. The correlation matrix table contains Pearson product-moment correlations of Y variables. Correlation measures the strength of the linear relationship between two variables.

E. The Shafer Reference

Shafer describes a scalable parallel classifier for data mining. A decision-tree-based classification algorithm, called SPRINT, removes all memory restrictions and is fast and scalable.

F. The Bridges Reference

Bridges describes an indexing system and method for improving retrieval of data based on a query from a user from a database management system including a main computer and a memory coupled to the main computer for storing the data. The indexing system includes a parallel computer coupled to the main computer and a parallel disk array coupled to the parallel computer. The invention includes the steps of storing record based data in the memory of the database management system, storing a value based index of selected attributes related to the record based data on the parallel disk array, and determining whether the parallel computer can be used to execute a query to obtain at least a partial result to the query. If so, the query is sent to the parallel computer and the query is executed on the parallel computer to obtain at least a partial result. If a final result cannot be determined on the parallel computer, the partial result from the parallel computer is sent to the database management system and a final result is determined on the database management system using the partial result received from the parallel computer.

G. Arguments Directed To The First Grounds for Rejection: Whether Claims 24-34, 36-40, 42-43, 46-55, 57-61, 63-74, 76-80, 82, and 83 Are Anticipated Under 35 U.S.C. §102(e) By Iyer et al., U.S. Patent No. 5,899,992 (Iyer).

1. Claims 24-34, 36-40 and 42-43

Appellants' attorney respectfully submits that Appellants' claims 24-34, 36-40, 42-43, 46-55, 57-61, 63-74, 76-80, 82, and 83 are patentable over the cited reference, because the reference does not teach or suggest the specific combination of elements recited in Appellants' independent claims.

Specifically, the reference does not teach or suggest the limitation of claim 24 directed to "an analytic application programming interface (API) that generates a set of scalable data mining functions including queries for execution by the relational database management system, executed by the computer, for performing data mining operations directly within the database management system."

The Office Action asserts that Iyer teaches these limitations of Appellants' claims at col. 3, line 50 through col. 4, line 26 (actually col. 4, line 35), which is set forth below:

> The scalable set-oriented classifier 114 of the present invention resorts to proven scalable database technology to provide a generic solution to the classification problem of scalability. The present invention provides a scalable model for classifying rows of a table within a classification tree. The scalable set-oriented classifier 114 is called the Scalable Supervised Learning Irregardless of Memory (SLIM) Classifier 114. Not only is the SLIM classifier 114 scalable in regions where recently published classifiers are not, but by virtue of building on well known set-oriented database management system (DBMS) primitives, the SLIM classifier 114 instantly exploits several decades of database research and development. The present invention rephrases classification, a data mining method, into analysis of data in a star schema, formalizing further the interrelationship between data mining and data warehousing.
>
> A description of a prototype built using IBM's DB2 product as the RDBMS 108, and experimental results for the prototype are discussed below. Generally, the experimental results indicate that the DB2-based SLIM classifier 114 has desirable properties associating it with linear scalability.
>
> The SLIM classifier 114 is built based on a set-oriented access to data paradigm. The SLIM classifier 114 uses Structured Query Language (SQL), offered by most commercial RDBMS 108 vendors, as the basis for the method. The SLIM classifier 114 is based on well known database methodologies and lets

the RDBMS 108 automatically handle scalability. As a result, the SLIM classifier 114 will scale as long as the database scales.

The SLIM classifier 114 leverages the Structured Query Language (SQL) Application Programming Interface (API) of the RDBMS 108, which exploits the benefits of many years research and development pertaining to:

(1) scalability
(2) memory hierarchy
(3) parallelism ([18])
(4) optimization of the executions([16])
(5) platform independence
(6) client server API ([17]).

See S. Sarawagi, Query Processing in Tertiary Memory Databases, VLDB 1995, [hereinafter Sarawag]; S. Sarawagi and M. Stonebraker, Benefits of Reordering Execution in Tertiary Memory Databases, VLDB 1996, [[hereinafter Stonebraker]; G. Bhargava, P. Goel, and B. Iyer, Hypergraph Based Reordering of Outer Join Queries with Complex Predicates, SIGMOD 1995, [hereinafter Bhargava]; T. Nguyen and V. Srinivasan, Accessing Relational Databases from the World Wide Web, SIGMOD 1996, [hereinafter Goel]; C. K. Baru et. al., DB2 Parallel Edition, IBM Systems Journal, Vol. 34, No 2, 1995, [hereinafter Baru]; each of which is which is incorporated by reference herein.

Appellants' attorney disagrees with this analysis.

The only API discussed in the above portions of Iyer is the Structured Query Language (SQL) Application Programming Interface (API) of the relational database management system (RDBMS). However, nothing in this description teaches or suggests "an analytic application programming interface (API) that generates a set of scalable data mining functions including queries for execution by the relational database management system, executed by the computer, for performing data mining operations directly within the database management system." Instead, this API of the RDBMS in Iyer only invokes functions of the RDBMS, but says nothing about generating a set of scalable data mining functions as recited in Appellants' claims. Moreover, the scalable set-oriented classifier of Iyer is not analogous to Appellants' claimed analytic API.

Indeed, the above portions of Iyer do not provide a proper basis for rejecting claim 24, because nowhere is Iyer properly applied to the limitations of claim 24. Instead, the Office Action relies on general conclusory statements regarding Iyer to reject Appellants' claims, without addressing the specific limitations of the claim or the specific teachings of Iyer.

Moreover, Appellants' claimed invention provides operational advantages over the system disclosed in Iyer. Moreover, Appellants' claimed invention solves problems not recognized by Iyer.

Thus, Appellants' attorney submits that independent claim 24 is allowable over Iyer. Further, dependent claims 25-34, 36-40 and 42-43 are submitted to be allowable over Iyer in the same manner, because they are dependent on independent claim 24, and thus contain all the limitations of the independent claim. Moreover, dependent claims 25-34, 36-40 and 42-43 recite additional novel elements not shown by Iyer. The arguments set forth below are directed not only to claims 25-34, 36-40 and 42-43, but to their grouped counterpart claims 46-55, 57-61, 63-74, 76-80, 82 and 83 as well.

In addition, dependent claims 46-55, 57-61, 63-74, 76-80, 82 and 83 are submitted to be allowable over Iyer in the same manner, because they are dependent on independent claims 44 and 45, and thus contain all the limitations of independent claims 44 and 45. Moreover, dependent claims 25-34, 36-40 and 42-43 recite additional novel elements not shown by Iyer.

2. Claims 25, 46 and 65

With regard to grouped dependent claims 25, 46 and 65, which recite that the computer comprises a parallel processing computer comprised of a plurality of nodes, and each node executes one or more threads of the relational database management system to provide parallelism in the data mining operations, the Office Action asserts that these limitations are taught by Iyer. Appellants' attorney disagrees. At the indicated location, Iyer refers to nodes of a classification tree, not the nodes of a parallel processing computer that executes threads of a relational database management system.

3. Claims 26, 47 and 66

With regard to grouped dependent claims 26, 47 and 66, which recite that the scalable data mining functions process data collections stored in the relational database and produce results that are stored in the relational database, these claims stand or fall with claims 24, 44 and 45, respectively.

4. Claims 27, 48 and 67

With regard to grouped dependent claims 27, 48 and 67, which recite that the scalable data mining functions are created by parameterizing and instantiating the analytic API, the Office Action asserts that these limitations are taught by Iyer. Appellants' attorney disagrees. At the indicated location, Iyer refers to an SQL API, not an analytic API for creating scalable data mining functions.

5. Claims 28, 49 and 68

With regard to grouped dependent claims 28, 49 and 68, which recite that the scalable data mining functions are dynamically generated queries comprised of combined phrases with substituting values therein based on parameters supplied to the analytic API, the Office Action asserts that these limitations are taught by Iyer. Appellants' attorney disagrees. At the indicated location, Iyer refers to an SQL API, not an analytic API for creating scalable data mining functions, and thus does not dynamically generate queries comprised of combined phrases with substituting values therein based on parameters supplied to the analytic API.

6. Claims 29, 50 and 69

With regard to grouped dependent claims 29, 50 and 69, which recite that the scalable data mining functions comprise Data Description functions, Data Derivation functions, Data Reduction functions, Data Reorganization functions, Data Sampling functions, or Data Partitioning functions, these claims stand or fall with claims 28, 49 and 68, respectively.

7. Claims 30, 51 and 70

With regard to grouped dependent claims 30, 51 and 70, which recite that the Data Description functions comprise descriptive statistical functions, these claims stand or fall with claims 29, 50 and 69, respectively.

8. Claims 31, 52 and 71

With regard to grouped dependent claims 31, 52 and 71, which recite that the Data Description functions comprise:

(1) descriptive statistics for one or more numeric columns, wherein the statistics are selected from a group comprising count, minimum, maximum, mean, standard deviation, standard mean error, variance, coefficient of variance, skewness, kurtosis, uncorrected sum of squares, corrected sum of squares, and quantiles,

(2) a count of values for a column,

(3) a calculated modality for a column,

(4) one or more bin numeric columns of counts with overlay and statistics options,

(5) one or more automatically sub-binned numeric columns giving additional counts and isolated frequently occurring individual values,

(6) a computed frequency of one or more column values,

(7) a computed frequency of values for pairs of columns in a column list,

(8) a Pearson Product-Moment Correlation matrix,

(9) a Covariance matrix,

(10) a sum of squares and cross-products matrix, or

(11) a count of overlapping column values in one or more combinations of tables,

these claims stand or fall with claims 29, 50 and 69, respectively.

<u>9. Claims 32, 53 and 72</u>

With regard to grouped dependent claims 32, 53 and 72, which recite that the Data Derivation functions provide column derivations or transformations, the Office Action asserts that these limitations are taught by Iyer. Appellants' attorney disagrees. At the indicated location, Iyer merely describes transforming table names, but says nothing about column derivations or transformations.

<u>10. Claims 33, 54 and 73</u>

With regard to grouped dependent claims 33, 54 and 73, which recite that the Data Derivation functions comprise:

(1) a derived binned numeric column wherein a new column is bin number,

(2) a n-valued categorical column dummy-coded into "n" 0/1 values,

(3) a n-valued categorical column recoded into n or less new values,

(4) one or more numeric columns scaled via range transformation,

(5) one or more columns scaled to a z-score that is a number of standard deviations from a mean,

(6) one or more numeric columns scaled via a sigmoidal transformation function,

(7) one or more numeric columns scaled via a base 10 logarithm function,

(8) one or more numeric columns scaled via a natural logarithm function,

(9) one or more numeric columns scaled via an exponential function,

(10) one or more numeric columns raised to a specified power,

(11) one or more numeric columns derived via user defined transformation function,

(12) one or more new columns derived by ranking one or more columns or expressions based on order,

(13) one or more new columns derived with quantile 0 to n-1 based on order and n,

(14) a cumulative sum of a value expression based on a sort expression,

(15) a moving average of a value expression based on a width and order,

(16) a moving sum of a value expression based on a width and order,

(17) a moving difference of a value expression based on a width and order,

(18) a moving linear regression value derived from an expression, width, and order,

(19) a multiple account/product ownership bitmap,

(20) a product ownership bitmap over multiple time periods,

(21) one or more counts, amount, percentage means and intensities derived from a transaction summary,

(22) one or more variabilities derived from transaction summary data,

(23) one or more derived trigonometric values and their inverses, including sin, arcsin, cos, arccos, csc, arccsc, sec, arcsec, tan, arctan, cot, and arccot, or

(24) one or more derived hyperbolic values and their inverses, including sinh, arcsinh, cosh, arccosh, csch, arccsch, sech, arcsech, tanh, arctanh, coth, and arccoth,

the Office Action asserts that the limitations of element (12) are taught by Iyer. Appellants' attorney disagrees. At the indicated location, Iyer merely describes forming groupings, not new columns derived by ranking one or more columns or expressions based on order.

11. Claims 34, 55 and 74

With regard to grouped dependent claims 34, 55 and 74, which recite that the Data Reduction functions provide matrix building operations to reduce the amount of data required for analytic algorithms, the Office Action asserts that these limitations are taught by Iyer. Appellants' attorney disagrees. At the indicated location, Iyer merely describes the use of a "count matrix", but says nothing about matrix building operations that reduce the amount of data required for analytic algorithms.

12. Claims 36, 57 and 76

With regard to grouped dependent claims 36, 57 and 76, which recite that the Data Reorganization functions provide an ability to reorganize data by joining or de-normalizing pre-processed results into a wide analytic data set, these claims stand or fall with claims 29, 50 and 69, respectively.

13. Claims 37, 58 and 77

With regard to grouped dependent claims 37, 58 and 77, which recite that the Data Reorganization functions comprise:

(1) create a de-normalized new table by removing one or more key columns, or

(2) join a plurality of tables or views into a combined result table,

these claims stand or fall with claims 29, 50 and 69, respectively.

14. Claims 38, 59 and 78

With regard to grouped dependent claims 38, 59 and 78, which recite that the Data Sampling function provides an ability to construct a new table containing a randomly selected subset of the rows in an existing table or view, these claims stand or fall with claims 29, 50 and 69, respectively.

15. Claims 39, 60 and 79

With regard to grouped dependent claims 39, 60 and 79, which recite that the Data Sample function selects one or more data samples of specified sizes from a table these claims stand or fall with claims 29, 50 and 69, respectively.

16. Claims 40, 61 and 80

With regard to grouped dependent claims 40, 61 and 80, which recite that the Data Partitioning function provides an ability to construct a new table containing at least one randomly selected subset of the rows in an existing table or view, wherein the subsets are mutually distinct but all-inclusive subsets of data, these claims stand or fall with claims 29, 50 and 69, respectively.

17. Claims 42, 63 and 82

With regard to grouped dependent claims 42, 63 and 82, which recite that results of the data mining operations are stored in the relational databases, these claims stand or fall with claims 24, 44 and 45, respectively.

18. Claims 43, 64 and 83

With regard to grouped dependent claims 43, 64 and 83, which recite that the relational database management system further comprises an analytical logical data model that stores metadata and processing results from the Scalable Data Mining Functions, the Office Action asserts that these limitations are taught by Iyer. Appellants' attorney disagrees. At the indicated location, Iyer says nothing about an analytical logical data model that stores metadata and processing results from the scalable data mining functions, but instead merely refers to a training set and leaf node list.

H. Arguments Directed To The Second Grounds for Rejection: Whether Claims 35, 56, and 75 Are Obvious Under 35 U.S.C. §103(a) Over Iyer In View Of SAS Institute Inc.

1. Claims 35, 56 and 75

With regard to grouped dependent claims 35, 56 and 75, which recite that the Data Reduction functions comprise:

(1) build one or more data reduction matrices from a group comprising: (i) a Pearson-Product Moment Correlations matrix; (ii) a Covariances matrix; and (iii) a Sum of Squares and Cross Products (SSCP) matrix,

(2) export a resultant matrix, or

(3) restart a matrix operation,

these claims stand or fall with claims 29, 50 and 69, respectively.

I. Arguments Directed To The Third Grounds for Rejection: Whether Claims 41, 62, and 81 Are Obvious Over Iyer In View Of SPRINT.

1. Claims 41, 62 and 81

With regard to grouped dependent claims 41, 62 and 81, which recite that the Data Partitioning function selects one or more data partitions from a table using a database internal hashing technique, these claims stand or fall with claims 29, 50 and 69, respectively.

J. Arguments Directed To The Fourth Grounds for Rejection: Whether Claims 44 and 45 Are Obvious Over Iyer In view Of Bridges

1. Claims 44 and 45

Appellants' attorney respectfully submits that Appellants' claims 44 and 45 are patentable over the cited references, because the references do not teach or suggest the specific combination of elements recited in Appellants' independent claims.

Specifically, the references do not teach or suggest the limitations of claims 44 and 45 directed to "an analytic application programming interface (API) that generates a set of scalable data mining functions including queries for execution by the relational database management

system, executed by the computer, for performing data mining operations directly within the database management system."

The Office Action fails to specifically address this limitation, but Appellants' attorney assumes that the Office Action intended a rejection similar to that made against claim 24, wherein the Office Action asserts that Iyer teaches these limitations of Appellants' claims at col. 3, line 50 through col. 4, line 26 (actually col. 4, line 35), which is set forth below:

> The scalable set-oriented classifier 114 of the present invention resorts to proven scalable database technology to provide a generic solution to the classification problem of scalability. The present invention provides a scalable model for classifying rows of a table within a classification tree. The scalable set-oriented classifier 114 is called the Scalable Supervised Learning Irregardless of Memory (SLIM) Classifier 114. Not only is the SLIM classifier 114 scalable in regions where recently published classifiers are not, but by virtue of building on well known set-oriented database management system (DBMS) primitives, the SLIM classifier 114 instantly exploits several decades of database research and development. The present invention rephrases classification, a data mining method, into analysis of data in a star schema, formalizing further the interrelationship between data mining and data warehousing.
>
> A description of a prototype built using IBM's DB2 product as the RDBMS 108, and experimental results for the prototype are discussed below. Generally, the experimental results indicate that the DB2-based SLIM classifier 114 has desirable properties associating it with linear scalability.
>
> The SLIM classifier 114 is built based on a set-oriented access to data paradigm. The SLIM classifier 114 uses Structured Query Language (SQL), offered by most commercial RDBMS 108 vendors, as the basis for the method. The SLIM classifier 114 is based on well known database methodologies and lets the RDBMS 108 automatically handle scalability. As a result, the SLIM classifier 114 will scale as long as the database scales.
>
> The SLIM classifier 114 leverages the Structured Query Language (SQL) Application Programming Interface (API) of the RDBMS 108, which exploits the benefits of many years research and development pertaining to:
>
> (1) scalability
> (2) memory hierarchy
> (3) parallelism ([18])
> (4) optimization of the executions([16])
> (5) platform independence
> (6) client server API ([17]).
>
> See S. Sarawagi, Query Processing in Tertiary Memory Databases, VLDB 1995, [hereinafter Sarawag]; S. Sarawagi and M. Stonebraker, Benefits of Reordering Execution in Tertiary Memory Databases, VLDB 1996, [[hereinafter

Stonebraker]; G. Bhargava, P. Goel, and B. Iyer, Hypergraph Based Reordering of Outer Join Queries with Complex Predicates, SIGMOD 1995, [hereinafter Bhargava]; T. Nguyen and V. Srinivasan, Accessing Relational Databases from the World Wide Web, SIGMOD 1996, [hereinafter Goel]; C. K. Baru et. al., DB2 Parallel Edition, IBM Systems Journal, Vol. 34, No 2, 1995, [hereinafter Baru]; each of which is which is incorporated by reference herein.

Appellants' attorney disagrees with this analysis.

The only API discussed in the above portions of Iyer is the Structured Query Language (SQL) Application Programming Interface (API) of the relational database management system (RDBMS). However, nothing in this description teaches or suggests "an analytic application programming interface (API) that generates a set of scalable data mining functions including queries for execution by the relational database management system, executed by the computer, for performing data mining operations directly within the database management system." Instead, this API of the RDBMS in Iyer only invokes functions of the RDBMS, but says nothing about generating a set of scalable data mining functions as recited in Appellants' claims. Moreover, the scalable set-oriented classifier of Iyer is not analogous to Appellants' claimed analytic API.

Indeed, the above portions of Iyer do not provide a proper basis for rejecting claims 44 and 45, because nowhere is Iyer properly applied to the limitations of claims 44 and 45. Instead, the Office Action relies on general conclusory statements regarding Iyer to reject Appellants' claims, without addressing the specific limitations of those claims or the specific teachings of Iyer.

Further, Bridges fails to overcome the deficiencies of Iyer. Recall that Bridges was cited only for a parallel computer system, and thus is not related to these limitations.

Appellants' claimed invention provides operational advantages over the system disclosed in the various references. Moreover, Appellants' claimed invention solves problems not recognized by the cited references.

Thus, Appellants' attorney submits that independent claims 44, and 45 are allowable over Iyer and Bridges.

IX. CONCLUSION

In light of the above arguments, Appellants respectfully submit that the cited references do not anticipate nor render obvious the claimed invention. More specifically, Appellants' claims recite novel physical features which patentably distinguish over any and all references under 35 U.S.C. §§ 102 and 103. As a result, a decision by the Board of Patent Appeals and Interferences reversing the Examiner and directing allowance of the pending claims in the subject application is respectfully solicited.

Respectfully submitted,

GATES & COOPER LLP
Attorneys for Applicants

Howard Hughes Center
6701 Center Drive West, Suite 1050
Los Angeles, California 90045
(310) 641-8797

Date: September 6, 2005

By: [signature]
Name: George H. Gates
Reg. No.: 33,500

GHG/bjs

APPENDIX

1-23. (CANCELED)

24. A computer-implemented system for performing data mining applications, comprising:

(a) a computer having one or more data storage devices connected thereto, wherein a relational database is stored on one or more of the data storage devices;

(b) a relational database management system, executed by the computer, for accessing the relational database stored on the data storage devices; and

(c) an analytic application programming interface (API) that generates a set of scalable data mining functions including queries for execution by the relational database management system, executed by the computer, for performing data mining operations directly within the database management system.

25. The system of claim 24 above, wherein the computer comprises a parallel processing computer comprised of a plurality of nodes, and each node executes one or more threads of the relational database management system to provide parallelism in the data mining operations.

26. The system of claim 24, wherein the scalable data mining functions process data collections stored in the relational database and produce results that are stored in the relational database.

27. The system of claim 24, wherein the scalable data mining functions are created by parameterizing and instantiating the analytic API.

28. The system of claim 24, wherein the scalable data mining functions are dynamically generated queries comprised of combined phrases with substituting values therein based on parameters supplied to the analytic API.

29. The system of claim 28, wherein the scalable data mining functions comprise Data Description functions, Data Derivation functions, Data Reduction functions, Data Reorganization functions, Data Sampling functions, or Data Partitioning functions.

30. The system of claim 29, wherein the Data Description functions comprise descriptive statistical functions.

31. The system of claim 29, wherein the Data Description functions comprise:

(1) descriptive statistics for one or more numeric columns, wherein the statistics are selected from a group comprising count, minimum, maximum, mean, standard deviation, standard mean error, variance, coefficient of variance, skewness, kurtosis, uncorrected sum of squares, corrected sum of squares, and quantiles,

(2) a count of values for a column,

(3) a calculated modality for a column,

(4) one or more bin numeric columns of counts with overlay and statistics options,

(5) one or more automatically sub-binned numeric columns giving additional counts and isolated frequently occurring individual values

(6) a computed frequency of one or more column values,

(7) a computed frequency of values for pairs of columns in a column list,

(8) a Pearson Product-Moment Correlation matrix,

(9) a Covariance matrix,

(10) a sum of squares and cross-products matrix, or

(11) a count of overlapping column values in one or more combinations of tables.

32. The system of claim 29, wherein the Data Derivation functions provide column derivations or transformations.

33. The system of claim 29, wherein the Data Derivation functions comprise:

(1) a derived binned numeric column wherein a new column is bin number,

(2) a n-valued categorical column dummy-coded into "n" 0/1 values,

(3) a n-valued categorical column recoded into n or less new values,

(4) one or more numeric columns scaled via range transformation,

(5) one or more columns scaled to a z-score that is a number of standard deviations from a mean,

(6) one or more numeric columns scaled via a sigmoidal transformation function,

(7) one or more numeric columns scaled via a base 10 logarithm function,

(8) one or more numeric columns scaled via a natural logarithm function,

(9) one or more numeric columns scaled via an exponential function,

(10) one or more numeric columns raised to a specified power,

(11) one or more numeric columns derived via user defined transformation function,

(12) one or more new columns derived by ranking one or more columns or expressions based on order,

(13) one or more new columns derived with quantile 0 to n-1 based on order and n,

(14) a cumulative sum of a value expression based on a sort expression,

(15) a moving average of a value expression based on a width and order,

(16) a moving sum of a value expression based on a width and order,

(17) a moving difference of a value expression based on a width and order,

(18) a moving linear regression value derived from an expression, width, and order,

(19) a multiple account/product ownership bitmap,

(20) a product ownership bitmap over multiple time periods,

(21) one or more counts, amount, percentage means and intensities derived from a transaction summary,

(22) one or more variabilities derived from transaction summary data,

(23) one or more derived trigonometric values and their inverses, including sin, arcsin, cos, arccos, csc, arccsc, sec, arcsec, tan, arctan, cot, and arccot, or

(24) one or more derived hyperbolic values and their inverses, including sinh, arcsinh, cosh, arccosh, csch, arccsch, sech, arcsech, tanh, arctanh, coth, and arccoth.

34. The system of claim 29, wherein the Data Reduction functions provide matrix building operations to reduce the amount of data required for analytic algorithms.

35. The system of claim 29, wherein the Data Reduction functions comprise:

(1) build one or more data reduction matrices from a group comprising: (i) a Pearson-Product Moment Correlations matrix; (ii) a Covariances matrix; and (iii) a Sum of Squares and Cross Products (SSCP) matrix,

(2) export a resultant matrix, or

(3) restart a matrix operation.

36. The system of claim 29, wherein the Data Reorganization functions provide an ability to reorganize data by joining or de-normalizing pre-processed results into a wide analytic data set.

37. The system of claim 29, wherein the Data Reorganization functions comprise:

(1) create a de-normalized new table by removing one or more key columns, or

(2) join a plurality of tables or views into a combined result table.

38. The system of claim 29, wherein the Data Sampling function provides an ability to construct a new table containing a randomly selected subset of the rows in an existing table or view.

39. The system of claim 29, wherein the Data Sampling function selects one or more data samples of specified sizes from a table.

40. The system of claim 29, wherein the Data Partitioning function provides an ability to construct a new table containing at least one randomly selected subset of the rows in an existing table or view, wherein the subsets are mutually distinct but all-inclusive subsets of data.

41. The system of claim 29, wherein the Data Partitioning function selects one or more data partitions from a table using a database internal hashing technique.

42. The system of claim 24, wherein results of the data mining operations are stored in the relational databases.

43. The system of claim 24, wherein the relational database management system further comprises an analytical logical data model that stores metadata and processing results from the Scalable Data Mining Functions.

44. A method for performing data mining applications, comprising:

(a) storing a relational database on one or more data storage devices connected to a computer;

(b) accessing the relational database stored on the data storage devices using a relational database management system; and

(c) executing an analytic application programming interface (API) that generates a set of scalable data mining functions including queries for execution by the relational database management system, for performing data mining operations directly within the database management system.

45. An article of manufacture comprising logic embodying a method for performing data mining applications, comprising:

(a) storing a relational database on one or more data storage devices connected to a computer;

(b) accessing the relational database stored on the data storage devices using a relational database management system; and

(c) executing an analytic application programming interface (API) that generates a set of scalable data mining functions including queries for execution by the relational database management system, for performing data mining operations directly within the database management system.

46. The method of claim 44 above, wherein the computer comprises a parallel processing computer comprised of a plurality of nodes, and each node executes one or more threads of the relational database management system to provide parallelism in the data mining operations.

47. The method of claim 44, wherein the scalable data mining functions process data collections stored in the relational database and produce results that are stored in the relational database.

48. The method of claim 44, wherein the scalable data mining functions are created by parameterizing and instantiating the analytic API.

49. The method of claim 44, wherein the scalable data mining functions are dynamically generated queries comprised of combined phrases with substituting values therein based on parameters supplied to the analytic API.

50. The method of claim 49, wherein the scalable data mining functions comprise Data Description functions, Data Derivation functions, Data Reduction functions, Data Reorganization functions, Data Sampling functions, and or Data Partitioning functions.

51. The method of claim 50, wherein the Data Description functions comprise descriptive statistical functions.

52. The method of claim 50, wherein the Data Description functions comprise:

(1) descriptive statistics for one or more numeric columns, wherein the statistics are selected from a group comprising count, minimum, maximum, mean, standard deviation, standard mean error, variance, coefficient of variance, skewness, kurtosis, uncorrected sum of squares, corrected sum of squares, and quantiles,

(2) a count of values for a column,

(3) a calculated modality for a column,

(4) one or more bin numeric columns of counts with overlay and statistics options,

(5) one or more automatically sub-binned numeric columns giving additional counts and isolated frequently occurring individual values

(6) a computed frequency of one or more column values,

(7) a computed frequency of values for pairs of columns in a column list,

(8) a Pearson Product-Moment Correlation matrix,

(9) a Covariance matrix,

(10) a sum of squares and cross-products matrix, or

(11) a count of overlapping column values in one or more combinations of tables.

53. The method of claim 50, wherein the Data Derivation functions provide column derivations or transformations.

54. The method of claim 50, wherein the Data Derivation functions comprise:

(1) a derived binned numeric column wherein a new column is bin number,

(2) a n-valued categorical column dummy-coded into "n" 0/1 values,

(3) a n-valued categorical column recoded into n or less new values,

(4) one or more numeric columns scaled via range transformation,

(5) one or more columns scaled to a z-score that is a number of standard deviations from a mean,

(6) one or more numeric columns scaled via a sigmoidal transformation function,

(7) one or more numeric columns scaled via a base 10 logarithm function,

(8) one or more numeric columns scaled via a natural logarithm function,

(9) one or more numeric columns scaled via an exponential function,

(10) one or more numeric columns raised to a specified power,

(11) one or more numeric columns derived via user defined transformation function,

(12) one or more new columns derived by ranking one or more columns or expressions based on order,

(13) one or more new columns derived with quantile 0 to n-1 based on order and n,

(14) a cumulative sum of a value expression based on a sort expression,

(15) a moving average of a value expression based on a width and order,

(16) a moving sum of a value expression based on a width and order,

(17) a moving difference of a value expression based on a width and order,

(18) a moving linear regression value derived from an expression, width, and order,

(19) a multiple account/product ownership bitmap,

(20) a product ownership bitmap over multiple time periods,

(21) one or more counts, amount, percentage means and intensities derived from a transaction summary,

(22) one or more variabilities derived from transaction summary data,

(23) one or more derived trigonometric values and their inverses, including sin, arcsin, cos, arccos, csc, arccsc, sec, arcsec, tan, arctan, cot, and arccot, or

(24) one or more derived hyperbolic values and their inverses, including sinh, arcsinh, cosh, arccosh, csch, arccsch, sech, arcsech, tanh, arctanh, coth, and arccoth.

55. The method of claim 50, wherein the Data Reduction functions provide matrix building operations to reduce the amount of data required for analytic algorithms.

56. The method of claim 50, wherein the Data Reduction functions comprise:

(1) build one or more data reduction matrices from a group comprising: (i) a Pearson-Product Moment Correlations matrix; (ii) a Covariances matrix; and (iii) a Sum of Squares and Cross Products (SSCP) matrix,

(2) export a resultant matrix, and or

(3) restart a matrix operation.

57. The method of claim 50, wherein the Data Reorganization functions provide an ability to reorganize data by joining or de-normalizing pre-processed results into a wide analytic data set.

58. The method of claim 50, wherein the Data Reorganization functions comprise:

(1) create a de-normalized new table by removing one or more key columns, or

(2) join a plurality of tables or views into a combined result table.

59. The method of claim 50, wherein the Data Sampling function provides an ability to construct a new table containing a randomly selected subset of the rows in an existing table or view.

60. The method of claim 50, wherein the Data Sampling function selects one or more data samples of specified sizes from a table.

61. The method of claim 50, wherein the Data Partitioning function provides an ability to construct a new table containing at least one randomly selected subset of the rows in an existing table or view, wherein the subsets are mutually distinct but all-inclusive subsets of data.

62. The method of claim 50, wherein the Data Partitioning function selects one or more data partitions from a table using a database internal hashing technique.

63. The method of claim 44, wherein results of the data mining operations are stored in the relational databases.

64. The method of claim 44, wherein the relational database management system further comprises an analytical logical data model that stores metadata and processing results from the Scalable Data Mining Functions.

65. The article of claim 45 above, wherein the computer comprises a parallel processing computer comprised of a plurality of nodes, and each node executes one or more threads of the relational database management system to provide parallelism in the data mining operations.

66. The article of claim 45, wherein the scalable data mining functions process data collections stored in the relational database and produce results that are stored in the relational database.

67. The article of claim 45, wherein the scalable data mining functions are created by parameterizing and instantiating the analytic API.

68. The article of claim 45, wherein the scalable data mining functions are dynamically generated queries comprised of combined phrases with substituting values therein based on parameters supplied to the analytic API.

69. The article of claim 68, wherein the scalable data mining functions comprise Data Description functions, Data Derivation functions, Data Reduction functions, Data Reorganization functions, Data Sampling functions, and or Data Partitioning functions.

70. The article of claim 69, wherein the Data Description functions comprise descriptive statistical functions.

71. The article of claim 69, wherein the Data Description functions comprise:

(1) descriptive statistics for one or more numeric columns, wherein the statistics are selected from a group comprising count, minimum, maximum, mean, standard deviation, standard mean error, variance, coefficient of variance, skewness, kurtosis, uncorrected sum of squares, corrected sum of squares, and quantiles,

(2) a count of values for a column,

(3) a calculated modality for a column,

(4) one or more bin numeric columns of counts with overlay and statistics options,

(5) one or more automatically sub-binned numeric columns giving additional counts and isolated frequently occurring individual values

(6) a computed frequency of one or more column values,

(7) a computed frequency of values for pairs of columns in a column list,

(8) a Pearson Product-Moment Correlation matrix,

(9) a Covariance matrix,

(10) a sum of squares and cross-products matrix, or

(11) a count of overlapping column values in one or more combinations of tables.

72. The article of claim 69, wherein the Data Derivation functions provide column derivations or transformations.

73. The article of claim 69, wherein the Data Derivation functions comprise:

(1) a derived binned numeric column wherein a new column is bin number,

(2) a n-valued categorical column dummy-coded into "n" 0/1 values,

(3) a n-valued categorical column recoded into n or less new values,

(4) one or more numeric columns scaled via range transformation,

(5) one or more columns scaled to a z-score that is a number of standard deviations from a mean,

(6) one or more numeric columns scaled via a sigmoidal transformation function,

(7) one or more numeric columns scaled via a base 10 logarithm function,

(8) one or more numeric columns scaled via a natural logarithm function,

(9) one or more numeric columns scaled via an exponential function,

(10) one or more numeric columns raised to a specified power,

(11) one or more numeric columns derived via user defined transformation function,

(12) one or more new columns derived by ranking one or more columns or expressions based on order,

(13) one or more new columns derived with quantile 0 to n-1 based on order and n,

(14) a cumulative sum of a value expression based on a sort expression,

(15) a moving average of a value expression based on a width and order,

(16) a moving sum of a value expression based on a width and order,

(17) a moving difference of a value expression based on a width and order,

(18) a moving linear regression value derived from an expression, width, and order,

(19) a multiple account/product ownership bitmap,

(20) a product ownership bitmap over multiple time periods,

(21) one or more counts, amount, percentage means and intensities derived from a transaction summary,

(22) one or more variabilities derived from transaction summary data,

(23) one or more derived trigonometric values and their inverses, including sin, arcsin, cos, arccos, csc, arccsc, sec, arcsec, tan, arctan, cot, and arccot, or

(24) one or more derived hyperbolic values and their inverses, including sinh, arcsinh, cosh, arccosh, csch, arccsch, sech, arcsech, tanh, arctanh, coth, and arccoth.

74. The article of claim 69, wherein the Data Reduction functions provide matrix building operations to reduce the amount of data required for analytic algorithms.

75. The article of claim 69, wherein the Data Reduction functions comprise:

(1) build one or more data reduction matrices from a group comprising: (i) a Pearson-Product Moment Correlations matrix; (ii) a Covariances matrix; and (iii) a Sum of Squares and Cross Products (SSCP) matrix,

(2) export a resultant matrix, and or

(3) restart a matrix operation.

76. The article of claim 69, wherein the Data Reorganization functions provide an ability to reorganize data by joining or de-normalizing pre-processed results into a wide analytic data set.

77. The article of claim 69, wherein the Data Reorganization functions comprise:

(1) create a de-normalized new table by removing one or more key columns, or

(2) join a plurality of tables or views into a combined result table.

78. The article of claim 69, wherein the Data Sampling function provides an ability to construct a new table containing a randomly selected subset of the rows in an existing table or view.

79. The article of claim 69, wherein the Data Sampling function selects one or more data samples of specified sizes from a table.

80. The article of claim 69, wherein the Data Partitioning function provides an ability to construct a new table containing at least one randomly selected subset of the rows in an existing table or view, wherein the subsets are mutually distinct but all-inclusive subsets of data.

81. The article of claim 69, wherein the Data Partitioning function selects one or more data partitions from a table using a database internal hashing technique.

82. The article of claim 45, wherein results of the data mining operations are stored in the relational databases.

83. The article of claim 45, wherein the relational database management system further comprises an analytical logical data model that stores metadata and processing results from the Scalable Data Mining Functions.